



Universität
Zürich UZH



UNIVERSITÄT LEIPZIG



SCHWEIZERISCHER NATIONALFONDS ZUR
FÖRDERUNG DER WISSENSCHAFTLICHEN FORSCHUNG

sms4science.ch

Seedmoney project „From speech to text – _Interlingual subtitles and subtitling practices in Switzerland“, Alexander Künzli (Geneva) / Elisabeth Stark (Zurich)
Premier atelier, Genève, First Workshop, Geneva, 17 Janvier 2019

Comment rendre ‚naturels‘ les sous-titres – la question de l’‘incomplétude‘ linguistique

Elisabeth Stark
estark@rom.uzh.ch



Plan

1. Introduction: la notion d'ellipse
2. Le rapport phonie - graphie
3. Nos projets sur 'l'écrit abrégé'
4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?
5. Discussion et conclusion

1. Introduction: la notion d'ellipse

Définition et description de l'ellipse dans la grammaticographie française (p.ex. Grevisse ¹³2001):

P. ex. « ellipse étymologique » - *à vrai dire, phénomène de variation entre phonie et graphie plus différents registres:*

- a. Omission du pronom impersonnel : « [...] dans la langue parlée, et surtout dans la langue populaire, le pronom impersonnel est souvent omis, particulièrement avec *y avoir* et *falloir*, mais aussi avec d'autres verbes. » (Grevisse ¹³2001: 315).
- b. Omission du *ne* de négation : « Dans la langue parlée, surtout dans la langue familière, le *ne* disparaît avec des fréquences variables : presque systématiquement au Québec, souvent à Paris et dans le Berry, beaucoup moins en Lorraine et en Wallonie. » (Grevisse ¹³2001: 1462).

1. Introduction: la notion d'ellipse

Définition et description de l'ellipse du point de vue linguistique:

“Ellipsis is a **mismatch** between sound and meaning in which **certain selectional requirements are not met** in the phonetic realization.” (Aelbrecht ppt: 7) - toujours *dépendance du contexte*, souvent aussi *contrôle structural* du contexte:

(1) *Il rentrera vers NNNN, moi _ quelques mn avant.*

(1)' *Il rentrera vers NNNN, moi ~~[je rentrerai]~~ quelques mn avant:*

➡ reconstruction certaine et non-ambigue des éléments omis possibles!

➡ Phénomène attesté dans le phonique et le graphique!

1. Introduction: la notion d'ellipse

Différents types d'ellipse:

- (2) *Les bronzés* (titre d'un film comique des années 90)
- (3) *Privé* (sur une porte dans un bar)
- (4) *Car aggravation nuit du ve a sa. retour me.* (SMS)
- (5) *Bon anniversaire!*
- (6) *Il boit.*
- (8) *Je me suis acheté un pull et mes enfants un jeu de cartes* (gapping)
- (9) *Pierre aime les fraises, et Marie aussi.* (stripping)
- (10) A: *Qui est arrivée en retard?* - B: *Pierre.*
- (11) *Pierre ne veut pas se marier maintenant, mail il devra.* (VP ellipsis /
Modal complement ellipsis)
- (12) *J'ai invité un garçon, mais je ne me rappelle plus qui.* (sluicing)

1. Introduction: la notion d'ellipse

Cas douteux: pas vraiment des ellipses, mais un cas ,d'incomplétude':

Argument drop:

1) Emploi absolu (non pas contrôlé par le contexte linguistique):

(13) *Oui, je vois.*

(14) *Vous ne pouvez pas le déranger maintenant, il mange.*

2) *Pro drop* – sujet nul/zéro:

(15) it.: *(io) canto* – fr: **(je) chante*

(16)' it.: *bisogna dirglielo* – fr. *(il) faut le lui dire* (cf. Grevisse!)

(17) *Ø Semaine prochaine, Ø m'exporte à Genève, je pense (WhatsApp)*

3) *Topic-drop*: (18)A: *Starwars mag ich sehr.* - B: *Ø Kenn' ich nicht.*

“These phenomena are not normally taken together with ellipsis, although they might very well be part of the same mechanism.” (Aelbrecht: 48) ₆



1. Introduction: la notion d'ellipse

Questions pour aujourd'hui:

1. Que savons-nous sur ,l'incomplétude' dans le ,graphique abrégé'?

2.a Différences irréductibles entre ,l'incomplétude phonique' et celle graphique?

2.b Utilisation possible des structures incomplètes graphiques aussi dans les sous-titres du phonique?

2. Le rapport phonie - graphie

Validité générale des sources graphiques pour la recherche linguistique?

Validité des recherches sur le ‘français électronique mobile’ pour notre projet
(= sous-titrage de l’oral) (question pertinente aussi et surtout pour la
linguistique historique)?

Rapport entre ‘écrit’ et ‘parlé’ (médium!) et le statut de la graphie ?

- 1) dépendance totale de la phonie – “miroir fidèle” du ‘code phonique’ (cf. p.ex. Martinet 1960, cf. sur la position “phonocentrique” aussi Béguelin 2012) ou
- 2) autonomie en tant que système sémiotique propre (cf. p.ex. Anis 1988)?
- 3) “*approches pondérées*” (selon Béguelin 2012), cf. p. ex. sur la différence typologique dans le marquage des informations morphosyntaxiques entre le code phonique et le code graphique, Blanche-Benveniste (2010 : 152-159).



2. Le rapport phonie - graphie

Validité des recherches sur le ‘français électronique mobile’ pour notre projet
(= sous-titrage de l’oral)?

Les SMS, messages WhatsApp etc. en tant que ‘graphies informelles’, ‘libérées’
de la norme orthographique

A. reflètent *partiellement* ce qui se passe dans le code phonique en
communication informelle (“français de l’immédiat”) –

B. mais aussi: des *conventions orthographiques* apprises à l’école!

C. et possèdent des **régularités morphosyntaxiques propres** et divergents
du standard (écrit):
ne reflètent pas (toujours) l’oral, mais sont peut-être reconnues / perçues
comme naturelles par les téléspectateurs?



3. Nos projets sur ,l'écrit abrégé'

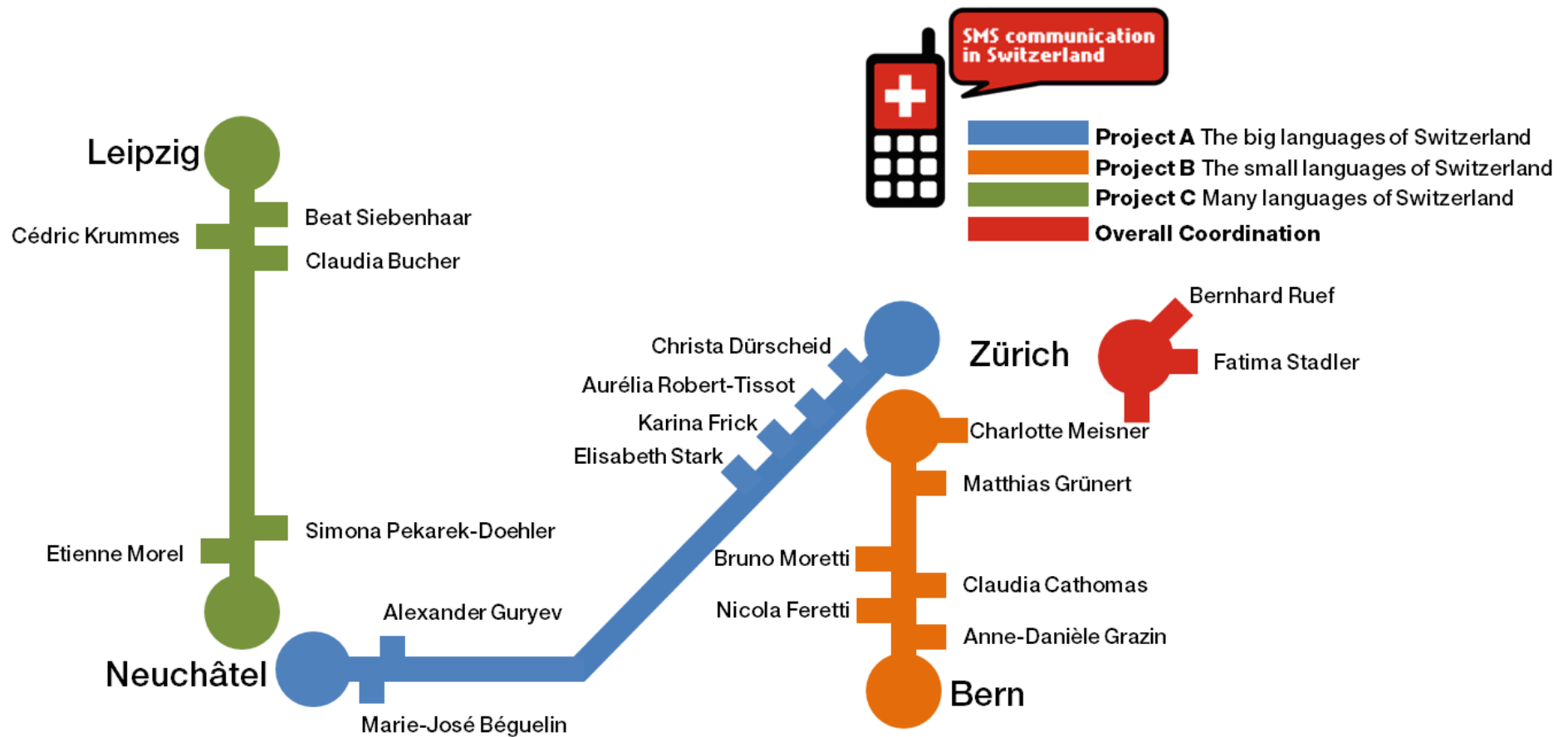
3.1 "SMS communication in Switzerland:

Facets of linguistic variation in a multilingual country" (FNS Sinergia; CRSII1_136230; 2011-2015; 7 doctorants, 3 post-docs, équipe TAL)

www.sms4science.ch:

- **Collecte:** du 01.09.2009 au 31.01.2010 (+ printemps 2011 pour une collecte additionnelle de données italophones et romanches)
- **SMS:** 25'947 (mots: ca. 500'000)
- **Participants:** 2'784
- **Sexe, âge, éducation etc. :** informations sociodémographiques disponibles d'environ 50% des participants (= 75% des SMS)
- **Langues/variétés principales:** dialectes alémanniques: 10'718 ; allemand non-dialectal: 7'224; français non-dialectal: 4'627; italien non-dialectal: 1'481, romanche: 1'121

“SMS Communication in Switzerland” - subprojects and partners



3. Nos projets sur ,l'écrit abrégé'

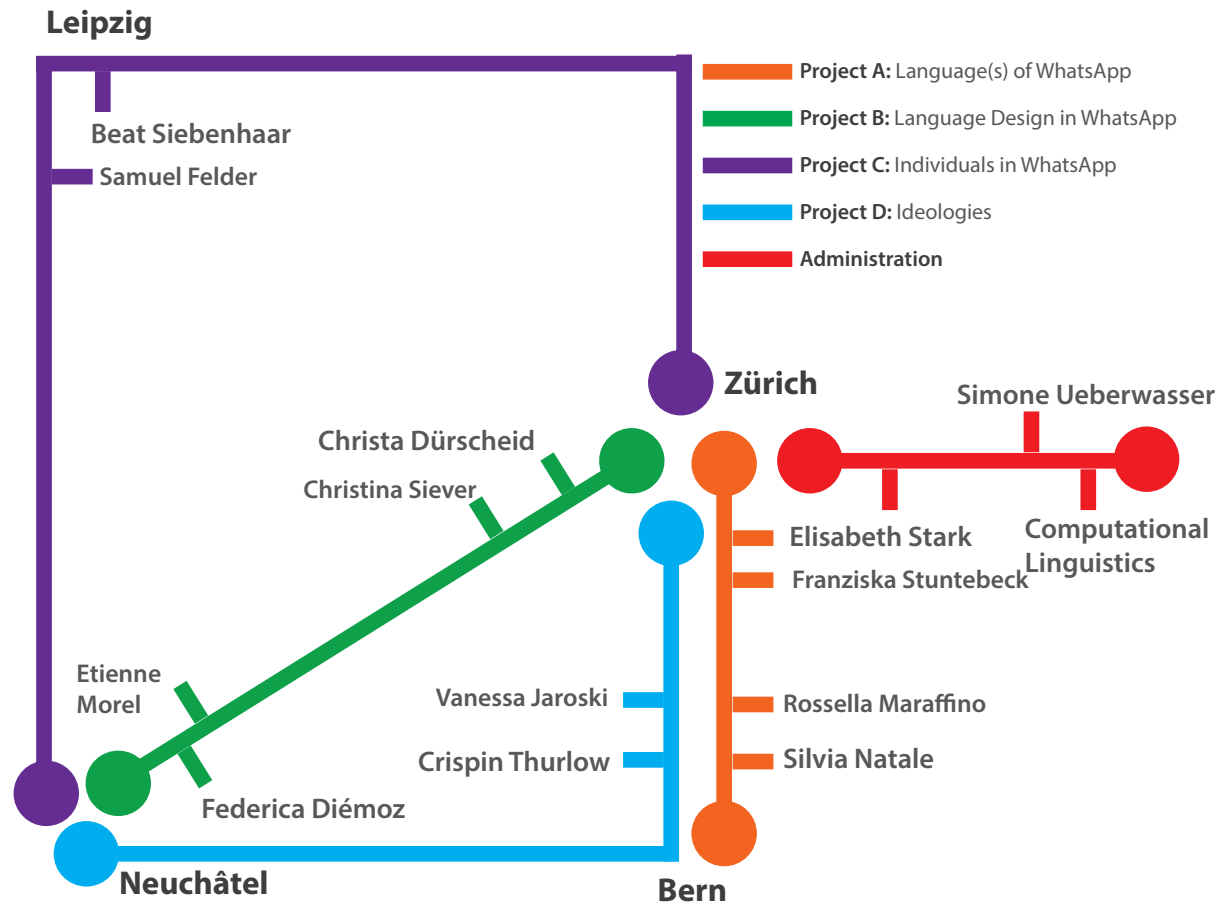
3.2 "What's up, Switzerland?"

"Language, Individuals and Ideologies in mobile messaging" (FNS Sinergia; CRSII1_160714; 2016-2020; 4 doctorants, 2 post-docs, équipe TAL)

- **Collecte:** été 2014
- **Chats :** ~617
- **Messages** (avec l'accord des participants) : ~750'000
- **Tokens :** ~5.5 Mio
- **Emojis :** ~350'000
- **Participants:** ~1000
- **Langues/variétés principales (messages):** dialectes alémaniques: 500'000 ; allemand non-dialectal: 81'000; français: 200'000; italien: 42'000, romanche: 29'000

www.whatsup-switzerland.ch

“What’s up, Switzerland?”



4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

4.1 Le marquage de la négation (cf. Stark 2012)

- Facteur le plus important: la catégorie grammaticale du sujet (cf. p.ex. chiffres comparable dans Hansen/Malderez 2004, Dufter/Stark 2007, van Compernelle 2008, 2010): Tableau 1:

	<i>-ne</i>	<i>+ne</i>	Σ
Sujet lexical	27 (34.2%)	52 (65.8%)	79 (100%)
Sujet pronominal	850 (80.2%)	210 (19.8%)	1060 (100%)

- (19) *Alors ma tenue **ne t'a pas** séduit hier?*
- (20) *Bah, **j'ai pas** encore trouvé un congrès où je me sentais [...] à ma place*
- (21) *Je **ne sais ps** si tu es au courant *ms Ø ns a demandé* [...]*

4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

Tableau 2: Type de sujet dans le corpus *sms4science.ch*:

Type	Exemple	-ne	%	+ne	%	Σ
A: clitic subjects	<i>je, tu, il ils, ce</i>	760	80.5	184	19.5	944
B. atypical clitic subjects	<i>elle, nous, vous, elles</i>	21	65.6	11	34.4	32
C: non-clitic subjects	<i>qui, ça, cela</i>	69	82.1	15	17.9	84
D: lexical subjects	<i>mon voisin</i>	27	34.2	52	65.8	79

(22) [...] *nous sommes allés y manger avec Rizvana et **nous** n'avons pas pu résister à prendre les ravioli Caren [...]*

4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

Tableau 3: Type de sujet dans le corpus *C-ORAL-ROM* (cf. Meisner 2010):

Type	Example	-ne	%	+ne	%	Σ
A: clitic subjects	<i>je, tu, il ils, ce</i>	1421	84	268	16	1689
B. atypical clitic subjects	<i>elle, nous, vous, elles</i>	125	65	66	35	191
C: non-clitic subjects	<i>qui, ça, cela</i>	145	55	121	45	266
D: lexical subjects	<i>mon voisin</i>	26	13	167	87	193

4. ‘L’incomplétude électronique graphique’ - miroir de l’oral ou régularités spécifiques?

4.2 L’ellipse du déterminant - purement graphique?!

Contextes ‘typiques’ d’omission de déterminant:

P.ex. “framing” – “cadres” (cf. Chafe 1976, Stark 1997) - aussi dans le code phonique?!

(23) *Sinon, Ø **bonne nouvelle**. J’ai au moins 5 à mon essai sur la poésie*
(23890)

(24) *Yo, comment va? Ø **Ptite question** boulot:jsui en train de faire du service civil et jvais demain faire remplir le questionnaire apg à mon employeur le +récent(le spo). Ensuite il vs l enverra. Tu penses que jaurai quand les sous sur mon compte?merci d’avance* (9191)

➔ “marqueurs de topic” (cf. NOAILLY-LE BIHAN 1982 sur *côté*, *question* etc. *question fric*, *ça va plutôt bien*).

4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

4.2 L'ellipse du déterminant - purement graphique?!

Contextes 'typiques' d'omission de déterminant:

P.ex. Ajouts et syntagmes prépositionnels

(25) *Parfait je suis au cinquieme gauche en sortant Ø Ø ascenseur*
(23373)

(26) *Ø Semaine prochaine, Ø m'exporte à Genève, je pense*

Omission du déterminant surtout dans des ajouts (prépositionnels) = positions non-argumentales - mais aussi des fois en position argumentale (sujet, COD):

(27) *Ø Pediatre dit ~~que~~ tt va tres bien* (17750)

(28) *Ø Tome 10 patois paraitra ce jour* (13081)

4. ‘L’incomplétude électronique graphique’ - miroir de l’oral ou régularités spécifiques?

4.3. L’ellipse du sujet

(29) *ø Suis au boulot...damage...tes sur ke vendredi ø yapud'place?*

- Phénomène noté depuis les années 80 du siècle dernier : “ subject drop in abbreviated registers ” (cf. numéro spécial de *Linguistic Variation* 17-2)
- Typique de certains ‘registres abrégés’ ou ‘simplifiés’ cf. *registers* comme “situationally defined varieties” selon Biber 1995: 1:
“Registers are language varieties used to communicate for specific purposes” (Paesani 2006: 149)
- Facteurs clé (cf. Biber 1995): intention communicative, *interaction*, conditions d’usage, relation entre interlocuteurs.

4. ‘L’incomplétude électronique graphique’ - miroir de l’oral ou régularités spécifiques?

4.3. L’ellipse du sujet

Idée: l’écrit électronique mobile est une forme de communication interactionnelle, mais aussi économique:

“reduced CP” (cf. Haegeman 1997, 1999, 2013, Weir 2012):

- Prédications de l’analyse de Haegeman:
pas de sujets omis dans les subordonnées, dans des structures interrogatives, en inversion, après des compléments préposés (mais possibles après des ajouts et après des conjonctions de coordination):



Ellipse spécifique, non comparable au *pro-drop* ou au *topic-drop* germanique - régularité purement graphique?



4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

4.3. L'ellipse du sujet

Quelques résultats du corpus SMS (Stark/Robert-Tissot 2017):

- Messages analysés: 1'100
- Nombre de sujets: 3'451 (dont 2'456 des sujets clitiques)
- Résultat global: 8.46% des sujets omis (9.4% des sujets et 9,7% des sujets clitiques dans le corpus SMS français entier:
Robert-Tissot 2018a: 248)

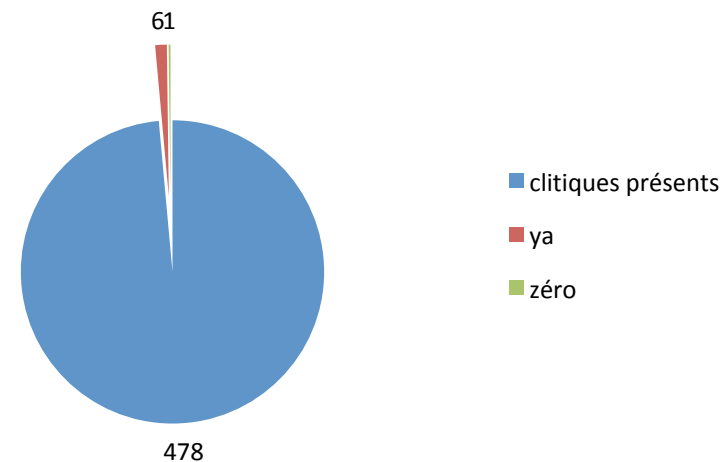
4. 'L'incomplétude électronique graphique' - miroir de l'oral ou régularités spécifiques?

4.3. L'ellipse du sujet

Quelques résultats du corpus SMS (Stark/Robert-Tissot 2017):

- Sur 478 sujets clitiques dans les subordonnées, seuls 7 sont omis, dont 6 dans *il y a* > *y a* et une faute de frappe (?) (cf. Stark/Robert-Tissot 2017: 262):

Zéro en subordonnées

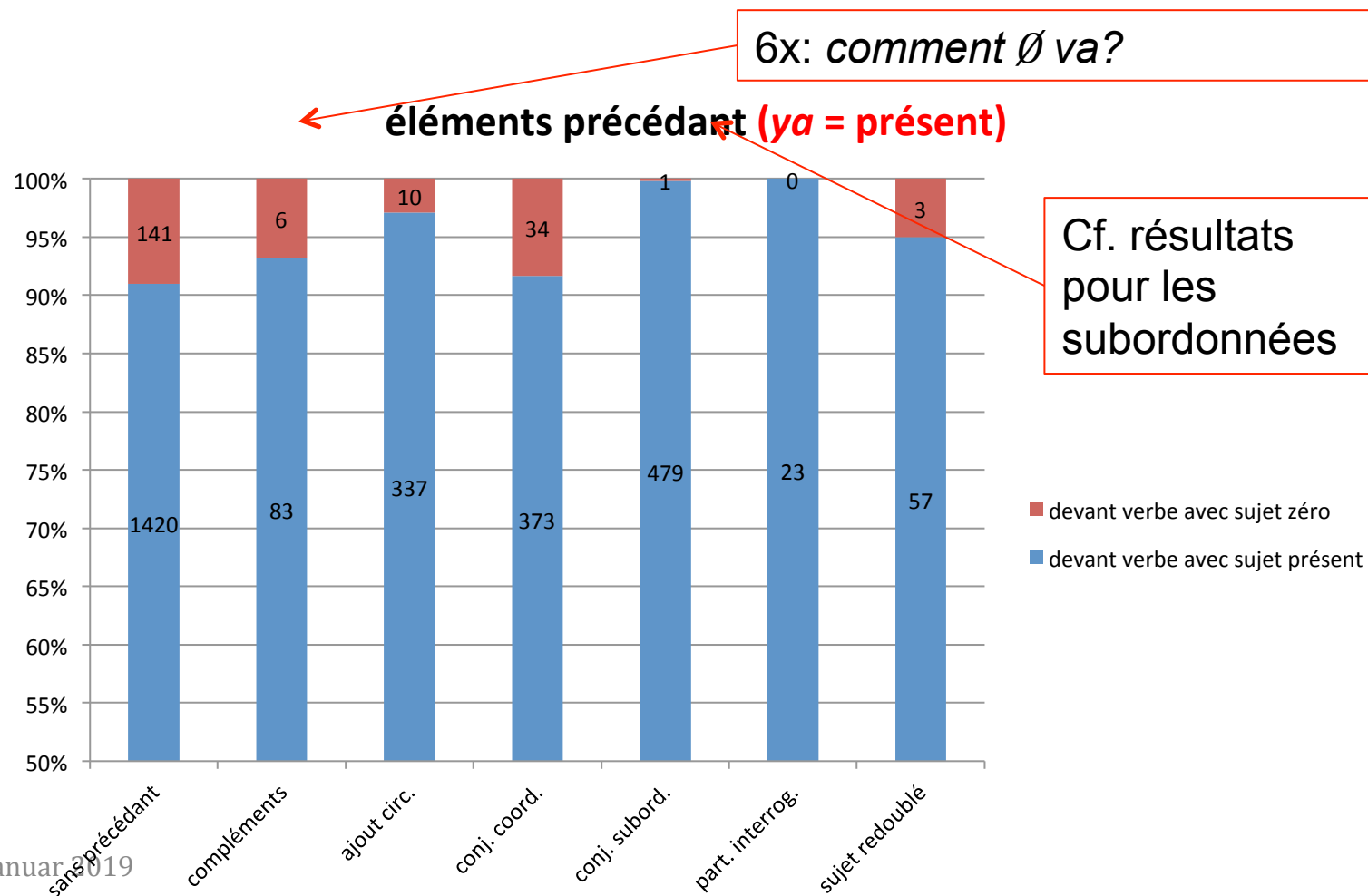


(30) *Mais tu peux juste me dire keske tu à fais ce week **pour ke Ø le puisse** déjà écrire stp!!bisou*

Donc: pas d'omission du sujet dans les subordonnées.

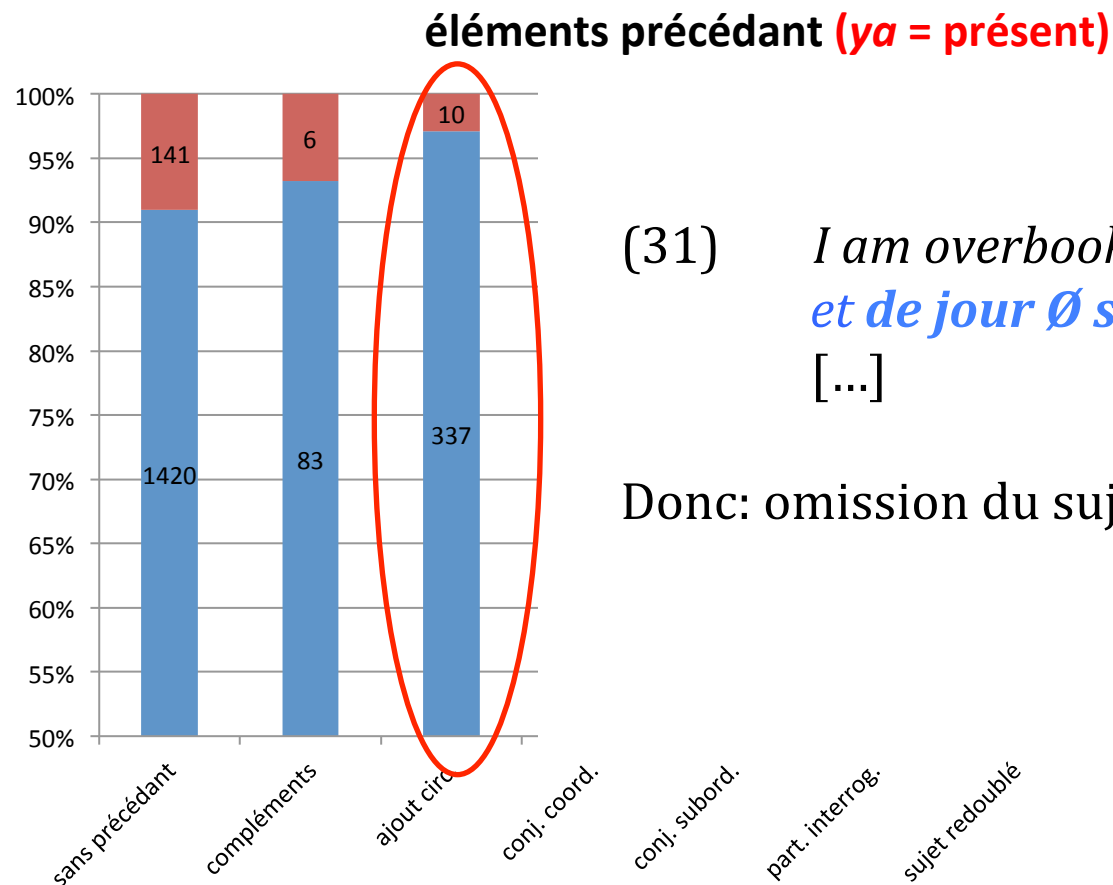
4.3. L'ellipse du sujet

Quelques résultats du corpus SMS (Stark/Robert-Tissot 2017; cf. aussi les résultats du corpus SMS français entier dans Robert-Tissot 2018a: 281):



4.3. L'ellipse du sujet

Quelques résultats du corpus SMS (Stark/Robert-Tissot 2017; cf. aussi les résultats du corpus SMS français entier dans Robert-Tissot 2018a: 281):



(31) *I am overbooked jusko vendredi soir:
et de jour Ø suis en stage à la radio!*
[...]

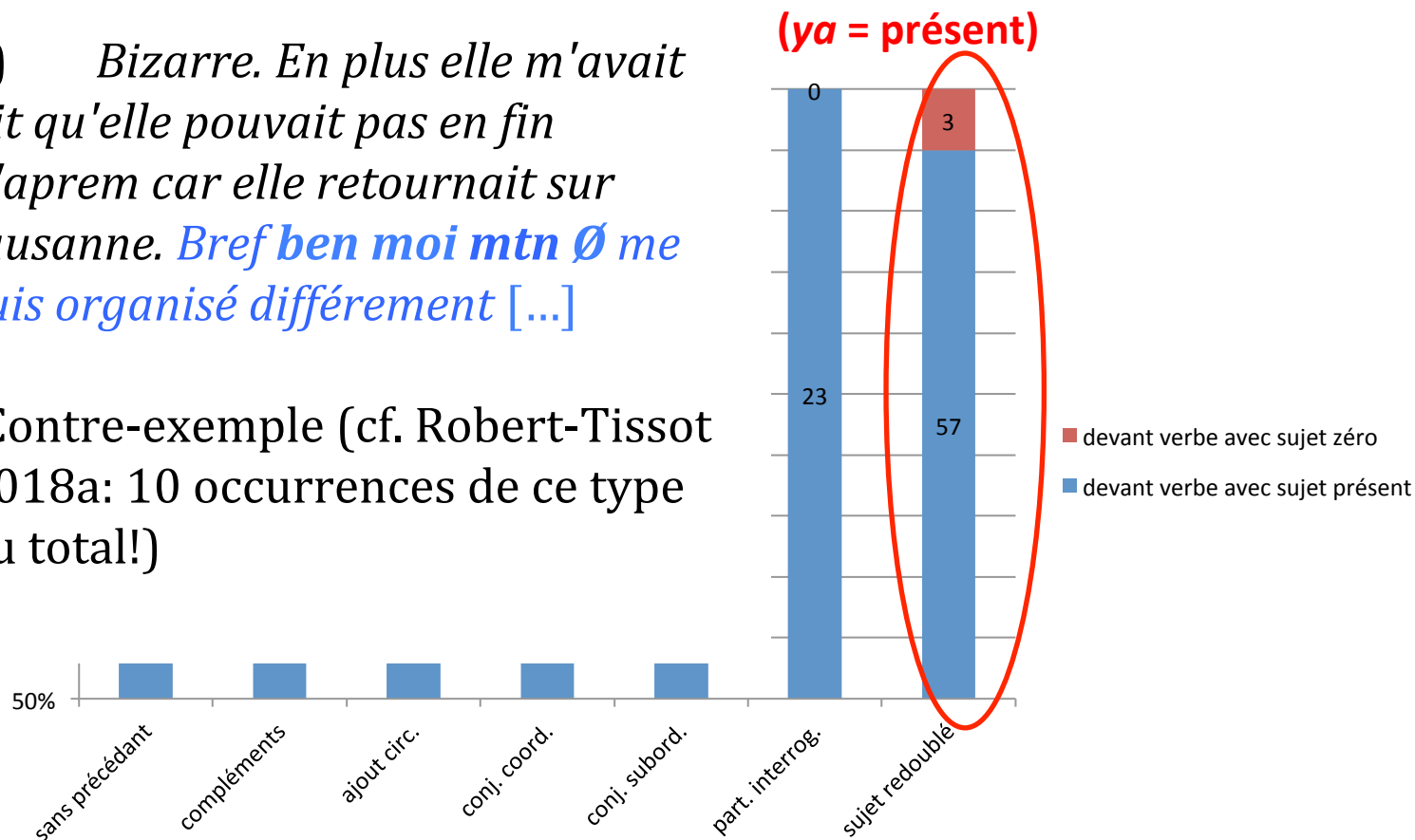
Donc: omission du sujet attestée après ajouts.

4.3. L'ellipse du sujet

Quelques résultats du corpus SMS (Stark/Robert-Tissot 2017; cf. aussi les résultats du corpus SMS français entier dans Robert-Tissot 2018a: 281):

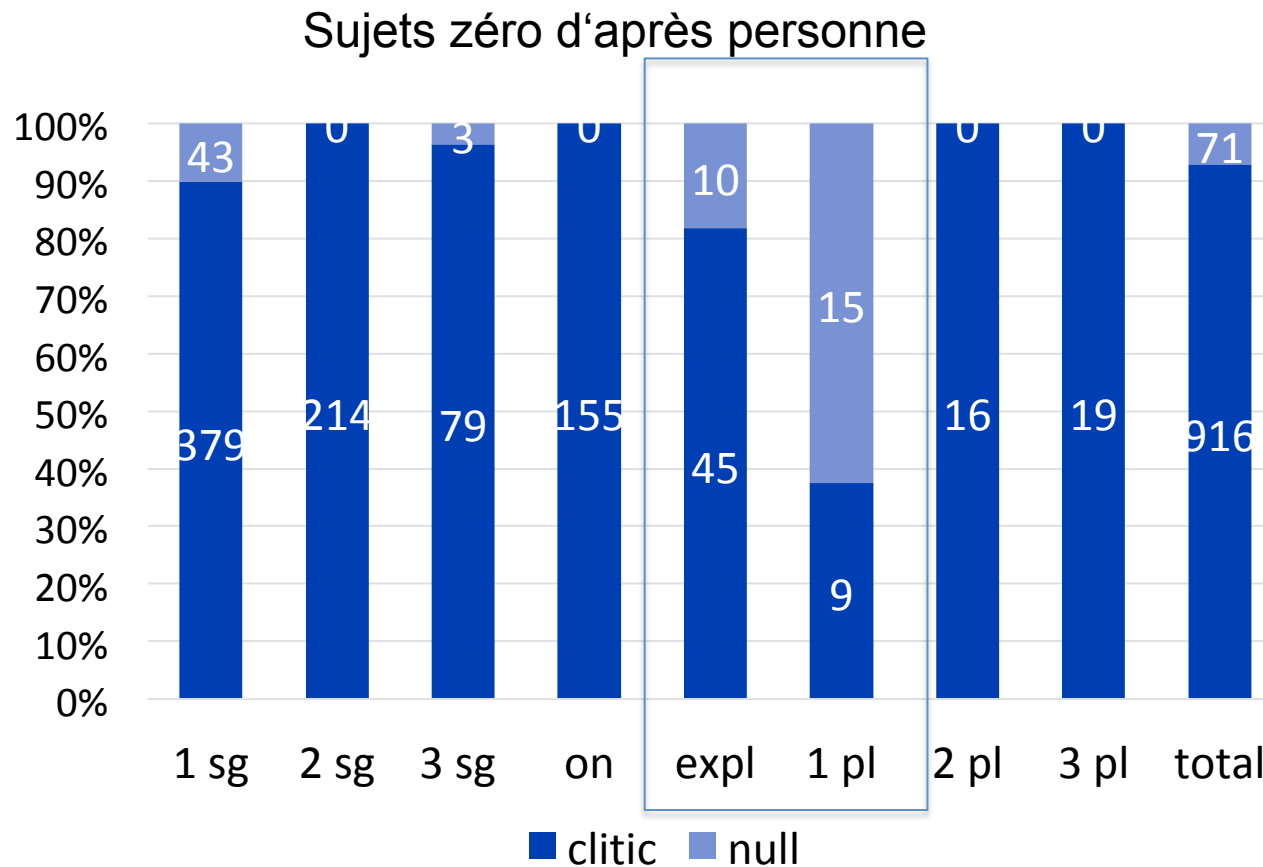
(32) *Bizarre. En plus elle m'avait dit qu'elle pouvait pas en fin d'aprem car elle retournait sur lausanne. Bref ben moi mtn Ø me suis organisé différemment [...]*

→ Contre-exemple (cf. Robert-Tissot 2018a: 10 occurrences de ce type au total!)



4.3. L'ellipse du sujet

Quelques résultats du corpus WhatsApp (Stuntebeck 2018, Stuntebeck en prép): résultats suivants se basent sur l'analyse de 4 chats / 8 personnes:



4.3. L'ellipse du sujet

Quelques résultats du corpus WhatsApp (Stuntebeck 2018, Stuntebeck en prép): résultats suivants se basent sur l'analyse de 4 chats / 8 personnes:

- (33) A: *J'espère que vous_i êtes bien rentrés. Nous on profite de notre dernière soirée ici. [...]* (531826)
- B: *Ø_i Sommes très bien rentrés. La bise s'est levée tout a l'heure et a chassé les nuages : beau temps mais très frisquet ! Ø_i Rentrons d'une expo chez mon collègue à genthod et Ø_i avons fait Ø route suisse magnifique. Bon retour bisou* (531827)



5. Discussion et conclusion

Rappel: Questions pour aujourd'hui:

1. Que savons-nous sur ,l'incomplétude' dans le ,graphique abrégé'?
- 2.a Différences irréductibles entre ,l'incomplétude phonique' et celle graphique?
- 2.b Utilisation possible des structures incomplètes graphiques aussi dans les sous-titres du phonique?

5. Discussion et conclusion

1. Que savons-nous sur ,l'incomplétude' dans le ,graphique abrégé'?

- Suit partiellement des régularités attestées aussi dans le phonique (omission du *ne* de négation), mais est souvent soumis à des régularités propres, p. ex. pour l'omission du sujet:
- observée dans d'autres types d'écrit abrégé' aussi (journaux intimes etc, cf. Haegeman/Stark en prép.), aussi en anglais, dans les ,contextes de Haegeman':

(34) *Jamais Ø ne se serait attendue à cela.*

(Journal intime de Léautaud 1933, 96, cf. Haegeman 2013: 91)

(35) *Ø Ai appris à ne pas tellement m'exposer à la flamme, et à courir peu de risques de loucher*

(36) *Ø Lendemain, Ø vais chez Martinet, Ø achète Ø manuel de dessin industriel*

(Journal intime de Simone Weil)

5. Discussion et conclusion

2.a Différences irréductibles entre ,l'incomplétude phonique' et celle graphique?

2.b Utilisation possible des structures incomplètes graphiques aussi dans les sous-titres du phonique?

- Régularités communes pour le *ne* de négation, le pronom impersonnel, l'omission des déterminants dans certaines constructions, mais aussi des régularités propres de l'“écrit abrégé” (omission du déterminant avec des arguments et ajouts (prépositionnels), omission du sujet référentiel (mais: exemple (37)?!):

(30) *Ø Suis au boulot...damage...tes sur ke vendredi Ø yapud'place?*

(37) *ouais d'ailleurs c'est ce qu'il me semblait bizarre au début quand euh j'ai entendu qu'elle faisait la colloc avec toi Ø me suis dit et ben | _ | _ | Ø me suis dit et ben c'est des filles qui ont rien à voir et puis qui se | _ | (OFROM, unifr12-jha, 215-287) - **corpus oral!** (Robert-Tissot 2018b)*



5. Discussion et conclusion

2.b Utilisation possible des structures incomplètes graphiques aussi dans les sous-titres du phonique?

?